

**IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF TENNESSEE
NASHVILLE DIVISION**

CONCORD MUSIC GROUP, INC., ET AL.,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case No. 3:23-cv-01092

Chief Judge Waverly D. Crenshaw, Jr.
Magistrate Judge Alistair Newbern

DECLARATION OF MICHAEL D. SMITH

I, **Michael D. Smith**, hereby declare pursuant to 28 U.S.C. § 1746:

1. I am the J. Erik Jonsson Professor of Information Technology and Marketing at Carnegie Mellon University's Heinz College of Information Systems Management and Public Policy Management.

2. I submit this declaration in connection with Plaintiffs' Reply in Support of Motion for Preliminary Injunction filed by Plaintiffs Concord Music Group, Inc., Capitol CMG, Inc., Universal Music Corp., Songs of Universal, Inc., Universal Music – MGB NA LLC, Polygram Publishing, Inc., Universal Music – Z Tunes LLC, and ABKCO Music, Inc. (collectively, "Publishers"), in their lawsuit against Defendant Anthropic PBC ("Anthropic").

3. My statements set forth below are based on my specialized knowledge, education, and experience, as applied to the facts and circumstances of this case. If called upon, I would and could competently testify as to the matters contained herein.

4. This declaration incorporates by reference my previous declaration, ECF No. 52, submitted in support of Plaintiffs' Motion for Preliminary Injunction, ECF Nos. 40–41. My

previous declaration contains important background and details not repeated here, including my curriculum vitae, areas of expertise, and initial conclusions relating to (1) the economic impact of Anthropic's use of Plaintiffs' copyrighted works on Plaintiffs and the songwriters they represent; and (2) the economic feasibility of complying with the preliminary injunction Plaintiffs seek on Anthropic.

5. In preparing this declaration, I relied on my general background and training and reviewed Publishers' filings, my previous declaration, Anthropic's Opposition to Plaintiffs' Motion for Preliminary Injunction ("Anthropic's Opposition"), ECF No. 67, and supporting filings, and other documents referenced herein.

6. The purpose and intent of this declaration is to provide a limited response to assertions made in Anthropic's Opposition.

I. Methodology, Evidence, and Analysis

7. The assertion made by Anthropic that my declaration "cites no data, studies, financial reports, or quantitative information at all"¹ is inaccurate to the extent it suggests I have not relied on analysis of empirical evidence in concluding Anthropic's use of Publishers' works in Claude's responses causes economic harm to Publishers.

8. My conclusions regarding economic harm follow directly from established economic theory and are informed by my years of training and experience as an economist with a particular focus on the impact of unlicensed distribution of copyrighted entertainment goods on the legal markets for those goods.

9. My conclusions are also supported by the overwhelming empirical evidence in the academic literature regarding the unlicensed use of copyrighted content, which reveal unlicensed

¹ Anthropic's Opposition at 16.

uses of copyrighted works have a causal effect on reduced revenue, investment, and the creation of fewer creative works.²

II. Harm to Publishers' Lyrics Display and Licensing Market

10. I next address Anthropic's assertion that Anthropic's use of Publishers' works do not affect the existing licensing market for Publishers' lyrics.

11. Anthropic and its experts, Dr. Peterson and Ms. Hall, do not challenge the existence of a licensing market for Publishers' lyrics in general. My initial declaration described the established licensing market for lyrics in detail. That existing market includes, among other things, licensing lyrics to online lyric aggregators and lyric websites, through which users can search and access lyrics online. This market is part of the overall market for Publishers' lyrics, as products of creative expression.

12. Rather than challenge the existence of the larger market for Publishers' lyrics or any particular component thereof, Anthropic's experts appear to conclude that Anthropic does not compete with Publishers for the use of their lyrics online.³ I disagree with this assessment for the following reasons.

13. First, Anthropic offers a service that directly competes with Publishers' licensees through Claude. The many instances of infringement cited by Publishers in their Complaint confirm that Claude is capable of identifying Publishers' works and displaying lyrics at a user's

² See, e.g., Brett Danaher, Michael D. Smith & Rahul Telang, *Piracy Landscape Study: Analysis of Existing and Emerging Research Relevant to Intellectual Property Rights Enforcement of Commercial-Scale Piracy* (U.S. Pat. and Trademark Off., Economic Working Paper No. 2020-02, March 20, 2020); Rahul Telang & Joel Waldfogel, *Piracy and New Product Creation: A Bollywood Story*, 43 INFO. ECON. AND POLICY, 1 (2018); Brett Danaher & Michael D. Smith, *Digital Piracy, Film Quality, and Social Welfare*, 24 GEO. MASON L. REV. 923 (2017).

³ See Anthropic's Opposition at 3–4; Declaration of Dawn R. Hall in Support of Defendant's Opposition to Plaintiffs' Motion for Preliminary Injunction ("Hall Decl."), ¶¶ 12, 38, ECF No. 67-17; Declaration of Steven R. Peterson, Ph.D. in Support of Defendant's Opposition to Plaintiffs' Motion for Preliminary Injunction, ("Peterson Decl.") ¶ 40, ECF No. 67-19.

request.⁴ Anthropic’s experts assert that it does not operate a “service[] that regurgitate[s] a wide array of song lyrics,”⁵ and therefore does not offer a competing service normally offered by Publishers or their licensees. But as demonstrated by the evidence submitted with Plaintiffs’ Complaint, Claude does precisely that.

14. Fundamental economics dictates that where two firms offer the same service to an audience, the two firms compete in the same market. Although Anthropic disavows its intent for users to utilize Claude “as a resource for users to obtain copies of already-existing materials in response to queries,”⁶ this runs contrary to the body of public statements made by Anthropic in recent months. For example, in a submission to the U.S. Copyright Office regarding artificial intelligence and copyright, Anthropic touted the many search capabilities and applications of Claude, claiming:

- “Claude tends to perform well at general, open-ended conversation; **search**, writing, editing, outlining, and summarizing text; coding; and providing helpful advice about a broad range of subjects.”⁷
- “Claude has been integrated into productivity tools offered by crowdsourced **question and answer platforms**, allowing users of those products to engage with a conversational assistant as they **search for information . . .**”⁸

⁴ Complaint, Exhibit A.

⁵ Peterson Decl. ¶ 40.

⁶ Declaration of Jared Kaplan in Support of Defendant’s Opposition to Plaintiffs’ Motion for Preliminary Injunction (“Kaplan Decl.”) ¶ 12, ECF No. 67-1. *See also* Anthropic’s Opposition at 4.

⁷ *See* Anthropic, Public Comments of Anthropic PBC on U.S. Copyright Office’s Notification of Inquiry and Request for Public Comments Regarding Artificial Intelligence and Copyright (Oct. 30, 2023) 2, https://downloads.regulations.gov/COLC-2023-0006-9021/attachment_1.pdf (emphasis added).

⁸ *Id.* at 3 (emphasis added).

- “Claude is also helping to power AI research assistants on AI-based **search engines** and chatbots.”⁹

15. Anthropic has also advertised Claude’s ability to “integrate seamlessly into web search”¹⁰ on its own website and announced efforts to “integrate Claude with sources of solid, real-time information like those found in search engines such as DuckDuckGo.”¹¹ This demonstrates to me Anthropic’s intent to participate in the online search market through Claude.

16. Anthropic’s finetuning process also indicates that Anthropic knew that users would use Claude to search for and deliver copyrighted lyrics. This is based on my review of datasets uploaded by Anthropic to the Hugging Face website that reflect Anthropic’s finetuning efforts to improve or refine Claude’s outputs.¹² Examples of prompts I saw in the dataset include:

- “What are the lyrics to American Pie by Don McLean?”¹³
- “I’m trying to figure out an artist by using the lyrics. Are you able to help?”¹⁴
- “Ok what song starts with the lyrics:”¹⁵
- “Please provide the lyrics for the song ‘Only Hope’ by Mandy Moore.”¹⁶

17. I understand the datasets reflect human feedback from Anthropic workers who initiated conversations with Claude and then chose preferred responses consistent with Anthropic’s

⁹ *Id.* at 4 (emphasis added).

¹⁰ *Put Claude to Work*, ANTHROPIC, <https://www.anthropic.com/product> (last visited Feb. 13, 2024)

¹¹ *Introducing Claude*, ANTHROPIC (Mar. 14, 2023), <https://www.anthropic.com/news/introducing-claude>.

¹² *Datasets: Anthropic/hh-rlhf*, HUGGING FACE, <https://huggingface.co/datasets/Anthropic/hh-rlhf>.

¹³ Exhibit B at 1.

¹⁴ *Id.* at 2.

¹⁵ *Id.* at 3.

¹⁶ *Id.* at 4.

objectives as part of the finetuning process.¹⁷ The documented use of finetuning prompts related to lyrics indicates to me that Anthropic understood that Claude could and would be used to search, identify, and display content including Publishers' lyrics as well as other copyrighted works when prompted. This is a service nearly identical to those offered by some of Publishers' licensees. Therefore, Anthropic operates as a competitor to Publishers' licensees in the general market for lyrics online, and its unauthorized usage and display of Publishers' lyrics harms the Publishers' existing market by eroding the market share of its licensees and weakening demand for licenses.

18. Second, Claude affects Publishers' existing markets by producing outputs that mimic or are derived from human works of authorship, including Publishers' lyrics. In addition to displaying copyrighted content, Claude is capable of, in the words of Anthropic, "developing more complex, creative outputs, such as novels and poetry."¹⁸ But where Claude generates creative outputs, such as lyrics, it produces products that compete with the works of human authors, such as songwriters and artists represented by Publishers and other rightsholders. This is particularly the case where Claude generates lyrics based on copyrighted works created by Publishers' songwriters. From an economic perspective, lyrics from Claude and human-made lyrics are competing products; they both offer creative expression. Simply put, by generating poems and musical lyrics to users upon request, Claude usurps demand for Publishers' lyrics as creative expression and usurps opportunities for Publishers' songwriters, especially where any outputs are based on creative works created by those very same songwriters.

¹⁷ Yuntao Bai et al., *Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback* (Apr. 12, 2022) 5, 65, arXiv:2204.05862v1, <https://arxiv.org/pdf/2204.05862.pdf>.

¹⁸ Kaplan Decl. ¶ 10.

19. Usurping demand for Publishers’ lyrics not only places Claude in competition with Publishers and the songwriters who create their lyrics but also creates downstream effects on Publishers’ licensing market, which is predicated on demand for the underlying lyrics. By siphoning away demand for lyrics, Anthropic also erodes Publishers’ ability to negotiate licenses for their lyrics.

20. Third, Anthropic’s participation in the market for Publishers’ lyrics without paying for the cost of goods erodes the entire value of the market. Where one firm in a market acquires an unfair advantage—such as not having to pay for inputs required to provide its services—its competitors are likely to lose market share and seek out similar competitive advantages to compensate. In other words, Anthropic’s failure to pay Publishers for the use of their lyrics allows Anthropic to display lyrics at a lower cost to consumers than those enterprises that duly acquire licenses. Contrary to Anthropic’s assertion, this indeed “threaten[s] [Publishers’] . . . licenses to existing lyric websites,”¹⁹ eroding licensees’ incentives to negotiate with Publishers and depressing the market for lyrics online.

21. The harm posed by displaying lyrics online without a license is a pronounced concern for Publishers’ licensees, as indicated by Genius’ recent lawsuit against Google and LyricFind for the display of lyrics lifted from Genius’ website by Google’s search engine,²⁰ which Genius alleged diverted significant web traffic and ad revenue.²¹ There, Genius characterized Google’s appropriation of Genius’ lyric transcriptions as placing Genius at “an unfair competitive disadvantage in the market for licensing and display of lyrics because [Google and LyricFind],

¹⁹ Peterson Decl. ¶ 40.

²⁰ Complaint, *Genius Media Group Inc. v. Google LLC et al.*, No. 1:19-cv-07279 (E.D.N.Y. Dec. 30, 2019), ECF No. 1.

²¹ *Id.* ¶¶ 100–04.

who have misappropriated content from Genius’ website, use[d] it to gain unfair advantages without the investment of time, effort and resources necessary to actually transcribe lyrics.”²² The same perceived unfair advantage is threatened here by Anthropic when it displays lyrics through Claude without obtaining licenses from Publishers and other rightsholders.²³

III. The Licensing Market for AI Training Materials

a. The Licensing Market for Song Lyrics

22. Anthropic’s assertion that “[t]here is no basis for the Court to find that a ‘burgeoning market’ for song lyrics as general-purpose AI training inputs . . . is likely to be developed in the future” is inconsistent with fundamental economic principles and market reality.²⁴

23. There is ample evidence that a licensing market for song lyrics as AI training inputs is developing or likely to develop. The basic economic definition of a market is “a collection of buyers and sellers that interact, resulting in the possibility for exchange.”²⁵ Such a market for licensed AI training data already exists, and Anthropic already participates in that market as a buyer/licensor. For example, Anthropic states that Claude’s training corpus includes datasets licensed from various “third-party businesses.”²⁶ Given that LLM licensing markets currently exist for other copyrighted works, there is no reason to believe that a market for lyrics could not similarly develop.

²² *Id.* ¶¶ 105–06.

²³ The Second Circuit upheld the dismissal of the case on the ground that Genius’ state law claims were preempted by the Copyright Act and Genius—as a nonexclusive licensee rather than a copyright owner—did not hold requisite rights to assert the claims. *See ML Genius Holdings LLC v. Google LLC*, 2022 WL 710744 (2d Cir. Mar. 10, 2022), *cert. denied*, 143 S. Ct. 2658 (2023). In my view, nothing in the court’s holding diminished the premise that Genius was economically harmed by a competitor displaying lyrics without paying for a license as other market participants do.

²⁴ Anthropic’s Opposition at 26. *See also* Peterson Decl. ¶ 6 (“LLMs would likely not exist if AI firms were required to license the works in their training datasets.”).

²⁵ ROBERT S. PINDYCK & DANIEL L. RUBINFELD, MICROECONOMICS 11 (2nd ed. 2006).

²⁶ Kaplan Decl. ¶ 22.

24. Dr. Peterson suggests that a market for licensing lyrics could not exist because the value of lyrics to LLMs is “small” relative to the high transaction costs,²⁷ citing OpenAI’s failed negotiations with the *New York Times* in support of the notion that the *Times* lacked “sufficient data to make it worthwhile for OpenAI to expend the cost and effort to negotiate a license to train.”²⁸ This line of reasoning arrives at the incorrect conclusion. That OpenAI spent time and resources to participate in discussions presupposes that *New York Times*’ content was valuable enough to OpenAI to enter discussions in the first place. Moreover, Dr. Peterson’s conclusion about why the negotiations failed is nothing more than speculation.

25. Publishers’ works also have value beyond serving as simple raw material to train LLMs to generate content, and may be used for internal research purposes that aid in the development of generative AI models. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

b. The Viability of a Sustainable Licensing Market

26. Anthropic’s statement that “the need for trillions of tokens of text to properly train general-purpose LLMs is incompatible with a working licensing market of sufficient scale . . . [and] the licensing market Plaintiffs posit simply could not exist for the inputs to technological tools like Claude” likewise misses the mark.³¹

²⁷ Peterson Decl. ¶ 14.

²⁸ *Id.* ¶ 30.

²⁹ See Exhibit A.

³⁰ Exhibit A at 20.

³¹ Anthropic’s Opposition at 26.

27. There is no evidence to suggest that a sustainable licensing market for training materials cannot meet the scale required to support the development of generative AI development.

28. On the cost side, Dr. Peterson's argument that transaction costs for licensing copyrighted content would be far too high to be sustainable is premised on the false notion that LLMs would inevitably have to negotiate with "millions of copyright owners of material on the internet" in order to develop functioning AI models.³² This fear is overstated. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] There is little reason to believe the necessary quantities of copyrighted content cannot be acquired by negotiated licenses. By way of example, as Ms. Hall notes in her declaration, lyrics service Genius has obtained licenses to millions of songs by "enter[ing] into licenses with every major publisher: Sony/ATV Music Publishing, EMI Music Publishing, Universal Music Publishing Group, and Warner/Chappell Music" in addition to "[developing] a form license with the National Music Publishers' Association (NMPA) which today covers more than 96% of the independent publisher market."³⁴ In the music streaming context, services like Amazon Music, Apple Music, and Spotify have also successfully negotiated licenses for millions of sound recordings from a whole spectrum of rightsholders.³⁵

³² Peterson Decl. ¶ 5(e).

³³ See Exhibit A.

³⁴ Hall Decl. ¶ 41, citing *Licensing*, GENIUS, <https://genius.com/static/licensing> (last visited Feb. 12, 2024). See also *infra*, ¶ 30.

³⁵ See *About Spotify*, SPOTIFY, <https://newsroom.spotify.com/company-info/> (last visited Feb. 12, 2024); Dave Johnson, *What is Amazon Music? Here's Everything you need to know*, AMAZON (Nov. 28, 2023), <https://www.aboutamazon.com/news/entertainment/what-is-amazon-music>; *Universal Music Group Expands Agreements with Amazon Music and Twitch*, LICENSE GLOBAL (Jan. 27, 2022), <https://www.licenseglobal.com/entertainment/universal-music-group-expands-agreements-amazon-music-and-twitch>; *Spotify and Universal Music Group Announce Global*,

29. Thus, the proper question to ask is whether a licensing market for Publishers' works (*i.e.*, lyrics) could exist with relatively low transaction costs. [REDACTED]

[REDACTED] The same can be done for musical works or lyrics. After all, as discussed above, sustainable large-scale licensing is familiar to and well established in the music industry.

30. I have also become aware of certifications offered to AI models that exclusively utilize licensed or copyright-free content.³⁷ Certified AI models include generative AI tools that utilize licensed content for training purposes.³⁸ This further exhibits an established and growing market for licensed training data between AI companies and rightsholders.

31. Licensing markets surrounding novel technologies like the ones at issue here often emerge where allowed to do so. For example, following its loss in *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001) related to copyright infringement arising out of its peer-to-peer file sharing network, Napster obtained licenses with major record groups to begin operating as a

Multi-Year License Agreement, UNIVERSAL MUSIC (July 22, 2020), <https://www.universalmusic.com/spotify-and-universal-music-group-announce-global-multi-year-license-agreement-2/>; Dani Deahl, *Spotify and Warner agree to an 'expanded' global licensing deal*, THE VERGE (Apr. 1, 2020), <https://www.theverge.com/2020/4/1/21203324/spotify-warner-agree-expanded-global-licensing-deal>; Jem Aswad, *Apple Music Signs New Licensing Deals with Majors (Report)*, VARIETY (Mar. 13, 2020), <https://variety.com/2020/digital/news/apple-music-new-licensing-deals-sony-universal-warner-music-1203533507/>.

³⁶ See Exhibit A.

³⁷ *Licensed Model Certification*, FAIRLY TRAINED, <https://www.fairlytrained.org/certifications> (last visited Feb. 12, 2024).

³⁸ See, e.g., *BRIA Legal Lobby*, BRIA, <https://bria.ai/terms-and-conditions/> (last visited Feb. 12, 2024) (“We train customizable models from scratch using our licensed datasets, optimized for enterprise needs.”).

legal music service.³⁹ The development of peer-to-peer technology similarly spurred record companies to “work[] with download sites and new subscription services to create a legitimate alternative to piracy networks” “all of which pay the creators.”⁴⁰

32. If LLMs like Anthropic’s AI model Claude were permitted to take lyrics and other copyrighted works without compensating rightsholders, this growing market for licensed copyrighted works, including Publishers’ works, would likely contract and may collapse. Likely or prospective participants would be discouraged from entering the market, while existing participants would be pushed out due to decreased negotiating leverage or higher costs of doing business compared to their competitors. But this would constitute a market failure, not a market feature. This is a scenario that has played out in a similar market at least once before where unlicensed use of Bollywood movies weakened market demand, and “[w]eaker effective demand undermined creative incentives” for Bollywood studios, resulting in reduced creative output and reduced creative quality.⁴¹

IV. Quantifying Economic Harms

33. Anthropic suggests harms to Publishers “are not ‘irreparable’ because [they] can be remedied by money damages” and “[b]ecause Plaintiffs routinely license the works at issue, there

³⁹ Jeff Leeds and Jon Healey, *Napster Strikes a Deal With 3 Major Labels*, LA TIMES (June 6, 2001), <https://www.latimes.com/archives/la-xpm-2001-jun-06-fi-6954-story.html>.

⁴⁰ *Piracy of Intellectual Property on Peer-to-Peer Networks: Hearing Before the Subcommittee on Courts, the Internet, and Intellectual Property of the Committee on the Judiciary*, 117th Cong. 54 (2002) (statement of Miranda Rosen, Chairman and Chief Executive Officer, Recording Industry Association of America), available at https://commdocs.house.gov/committees/judiciary/hju81896.000/hju81896_0.HTM.

⁴¹ Rahul Telang & Joel Waldfogel, *Piracy and New Product Creation: A Bollywood Story*, 43 INFO. ECON. AND POLICY, 1 (2018).

is no valid reason to assume that a license fee would not adequately compensate them for any alleged infringement by Anthropic.”⁴² Those statements are incorrect, for several reasons.

34. As a threshold matter, Dr. Peterson’s suggestion that existing licensing markets cannot accommodate LLM development, described above, is directly at odds with Anthropic’s argument that existing publishing licenses provide a straightforward path for determining any damages arising out of unauthorized uses of Publishers’ works. It is also at odds with the fact that Anthropic has engaged in some licensing negotiations for training data.

35. More importantly, it is difficult to quantify the exact effect of Anthropic’s use of Publishers’ works on future revenue and licensing opportunities. Causally identifying the precise amount of monetary harm presents a considerable econometric problem—one that goes far beyond Ms. Hall’s suggestion of using “standard economic models” and “comparable [licensing] agreements.”⁴³

36. First, any existing lyrics licensing agreements between Publishers and licensees for the use of their lyrics online were not negotiated with the developers of LLMs. By definition, the licenses relate only to the display of complete, authentic lyrics. They do not encompass the display of portions of Publishers’ works combined with other text, and certainly not for use as training data. Therefore, it is not clear how such agreements can be used as a “‘starting point or an aid’ in calculating licensing fee[s] or monetary damages” in the LLM setting.⁴⁴ Second, measuring the causal impact of LLM use on existing licensing markets is only part of the problem. Quantifying

⁴² Anthropic’s Opposition at 18. I also note Ms. Hall’s suggestion that “Publishers themselves recognize use of their Works can be quantified in monetary terms” mischaracterizes my prior statements. Hall Decl. ¶ 16. I stated only that Publishers’ lyrics have value, not that it is easy to quantify the lost value from unauthorized use of their lyrics in LLMs. Declaration of Michael D. Smith, ¶ 29, ECF No. 52.

⁴³ Hall Decl. ¶¶ 14–15.

⁴⁴ *Id.* ¶ 17.

the exact economic losses resulting from any reputational harms or loss of goodwill associated with Anthropic's unauthorized uses of Publishers' Works is an exercise in predicting the future without hope of true accuracy.

V. Excluding Works from the Training Datasets is Feasible

37. Anthropic has offered no evidence or facts to support its assertion that carving out copyrighted works from its training data on a prospective basis will be unfeasible. On the contrary, there are ample reasons to believe Anthropic is economically and technologically well-positioned to omit Publishers' copyrighted content from training datasets on a prospective basis.

38. First, Anthropic states that Publishers' lyrics make up only an "infinitesimally small fraction" of the overall training data set.⁴⁵ This stands to reason that excluding Publishers' Works would not have a significant effect on the size of the overall dataset.

39. Second, in connection with the preliminary injunction, Publishers have asked only for Anthropic to refrain from using Publishers' lyrics in future training and solely on a prospective basis. The preliminary injunction sought by Publishers would not require Anthropic to undo current training or extract data from currently utilized training sets, only future ones. Accordingly, Anthropic's estimated cost of [REDACTED] to retrain the current model in development does not reflect the true marginal cost of excluding Publishers' copyrighted content in future iterations of Claude. Rather, the marginal cost of prospectively excluding Publishers' copyrighted content from Anthropic's training corpus is simply the cost of identifying and removing that content from the corpus prior to the next time the model is retrained—retraining that Anthropic would have done anyway during their normal course of business. It is beyond doubtful that the marginal cost of

⁴⁵ Anthropic's Opposition at 2.

excluding of Publishers' lyrics on a prospective basis will render Anthropic's business—reportedly valued at \$18.4 billion⁴⁶—extinct.⁴⁷

40. Third, I have come to understand that removing certain datasets or websites from generative AI training sets is not only technologically feasible, but something other AI companies, including OpenAI, already do.⁴⁸ Specifically, OpenAI has claimed it “engaged in numerous productive dialogues with rightsholders, including authors and music publishers, and asked them to identify sites on the internet that reproduce their copyrighted works . . . [and] then been able to exclude those sites from being further crawled for future training,” in addition to “identi[fy]ing sites that have been identified by rightsholders as hosting infringing content in order to exclude those from being crawled for future training, as well.”⁴⁹ If OpenAI can do it, one would rationally presume that Anthropic can as well.

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct to the best of my personal knowledge and belief.

Executed in Pittsburgh, PA, this 14th day of February, 2024.



Michael D. Smith

⁴⁶ *Anthropic seeking to raise \$750 mln in funding round led by Menlo Ventures*, REUTERS (Dec. 20, 2023), <https://www.reuters.com/markets/deals/anthropic-raise-750-mln-menlo-ventures-led-funding-round-information-2023-12-21/>.

⁴⁷ For comparison, the estimated global market size of the entire music publishing industry is reportedly close to \$7.3 billion. Music Publishing Market Size, MORDOR INTELLIGENCE, <https://www.mordorintelligence.com/industry-reports/music-publishing-market/market-size> (last visited Feb. 2, 2024).

⁴⁸ See OpenAI, Public Comments of OpenAI on U.S. Copyright Office's Notification of Inquiry and Request for Comments Regarding Artificial Intelligence and Copyright (Oct. 30, 2023) at 10, https://downloads.regulations.gov/COLC-2023-0006-8906/attachment_1.pdf. See also Declaration of Ben Y. Zhao, ¶¶ 15–17; 27, ECF No. 47.

⁴⁹ Public Comments of Open AI, at 7.